Research Article

Snezana Savoska* and Blagoj Ristevski

Towards Implementation of Big Data Concepts in a Pharmaceutical Company

https://doi.org/10.1515/comp-2020-0201
Received Jul 31, 2019; accepted Jun 22, 2020

Abstract: Nowadays, big data is a widely utilized concept that has been spreading quickly in almost every domain. For pharmaceutical companies, using this concept is a challenging task because of the permanent pressure and business demands created through the legal requirements, research demands and standardization that have to be adopted. These legal and standards' demands are associated with human healthcare safety and drug control that demands continuous and deep data analysis. Companies update their procedures to the particular laws, standards, market demands and regulations all the time by using contemporary information technology. This paper highlights some important aspects of the experience and change methodology used in one Macedonian pharmaceutical company, which has employed information technology solutions that successfully tackle legal and business pressures when dealing with a large amount of data. We used a holistic view and deliverables analysis methodology to gain top-down insights into the possibilities of big data analytics. Also, structured interviews with the company's managers were used for information collection and proactive methodology with workshops was used in data integration toward the implementation of big data concepts. The paper emphasizes the information and knowledge used in this domain to improve awareness for the needs of big data analysis to achieve a competitive advantage. The main results are focused on systematizing the whole company's data, information and knowledge and propose a solution that integrates big data to support managers' decision-making processes.

Keywords: Big Data, Information Systems, Unstructured Data, Enterprise Resource Planning Systems, Serialization

***Corresponding Author: Snezana Savoska:** "St. Kliment Ohridski" University-Bitola, Faculty of Information and Communication Technologies – Bitola, Republic of North Macedonia; Email: snezana.savoska@uklo.edu.mk
Blagoj Ristevski: "St. Kliment Ohridski" University-Bitola, Faculty of Information and Communication Technologies – Bitola, Republic of North Macedonia; Email: blagoj.ristevski@uklo.edu.mk

1 Introduction

Faced with the challenges of globalization and market competition, some companies in the Republic of North Macedonia are turning towards using contemporary Information Technology (IT), which solves many problems and brings a lot of benefits to their business. But not all of them set their business priorities on the use of the emerging IT solutions. Only those companies that are globalized through foreign investments with share capital and some inventive and visionary companies successfully move forward in the new Internet economy era [30]. Moreover, the pharmaceutical companies were faced with a big pressure to adapt and follow the European Union (EU) [47] and national law regulations [46]. They use IT as a strategic tool that helps them to cope with this pressure [19]. But, only one Macedonian pharmaceutical company has managed to deal with the globalization, standardization and regulation issues through using recent IT and deployment of cutting-edge software tools that support big data concepts. The company has made improvements in its business processes by using emerging software tools chosen carefully to support EU demands. Hence, the company has established standards in the industry, market and customers' demands that have resulted in gaining a competitive advantage and increasing market shares [30].

The strategy of the company was to use emerging IT to handle the above-mentioned problems making specific business processes reengineering activities toward the Internet economy and quality demands. There were many separate applications as "data islands" in the company that were used to help in solving some arising ad-hoc problems and other issues on demand at that time. The expenses for IT were increasing over time, while the information quality was decreased. Also, the company was obliged to manage a huge amount of structured, semi-structured and unstructured data created in many business processes and legal demands, but unsuccessfully.

Since 2003, the company steering committee and top-level managers decided to make changes in the IT concepts, software architecture and methodology. These changes were towards searching for emerging IT solutions

that were supposed to be the "problem solvers" and the most suitable for the company's needs. They intended to make an integration of all software solutions and services in the company and to start with the implementation of enterprise resource planning (ERP) software tools, according to the leading software developers' recommendations. Using the proposed teamwork methodology, they implemented many appropriate software tools and integrated all data collected from various distributed databases in an SAP's company ERP system. They started to use a wide range of data generated in new processes as a result of the process reengineering. These data have big data characteristics: variety, velocity, variability, veracity, value and volume.

In this article, we describe how the company has responded to the legal and business demands by implementing a new application as strategic tools to solve the raising company's problems as well as systematized the main activities and used data, information and knowledge and propose some concepts toward using big data for decision-making. These efforts also led to increasing the market shares for the company [30], bringing many solutions to the problems and going toward using data for gaining helpful information following the big data concepts. Moreover, this implementation was moving towards creating business analytics and visualization tools as managers' decision support tools. The applied methodology is proactive, with supportive management roles, and it has created the prerequisites for big data analysis in joint workshops. The proposed solution is toward data integration and the creation of analytics tools to support company managers. We also propose directions for suitable improvement of the selected solutions for big data analysis.

The remainder of the paper is organized as follows. Related works for the big data concepts, using big data in pharmacy and related issues are described in the second section. Section 3 explains experiences and methodology that deals with problems, taking into account used technologies, resources, solutions and their benefits. The next section highlights the planned activities that emerge as new big data analysis and business challenges. The final section provides concluding remarks and offers some proposals and directions for further work.

2 Related works

Many studies are devoted to the big data concept, presenting main concepts and methods, technology and the latest insights in Business Intelligence and Analytics (BI&A) [10, 29]. They also consider the key big data properties: volume, value, variability, veracity, variety and velocity, denoted as 6 V's [7, 11, 17, 31]. The volume, which refers to the amount of data, is expected to rise drastically in the years ahead, and it is usually measured in terabytes, petabytes, or even yottabytes. Velocity alludes to data in motion, the speed and frequency of data creation, their processing and analysis. Heterogeneity and complexity of numerous datasets, which can be structured, semi-structured and unstructured, are related to the data variety. Veracity refers to the data quality, relevance, uncertainty, reliability and predictive value, while variability is related to the data consistency over time [31]. The value of the big data highlights the data analysis, making big data valuable for the data analysts and decision-makers in all organizations and companies that employ big data concepts.

During big data analysis, data quality through their intrinsic and contextual level and their attributes should be anticipated and appropriate methods for monitoring and controlling data quality should be applied [13]. There is a close association between the corporate capacity and data quality through data consistency and completeness [14]. One of the most important challenges of using big data in industry and manufacturing is continuous process improvement. It requires new insights in quality control and optimization using data-driven clustering in the processes of discovering real-time unusual patterns in the processes and analysis of root cause processes [24].

Big data acceptance in organizations can be made in diverse manners because of its ability to provide valuable benefits for organizations. Some researchers have proposed the use of the big data strategic grid. According to McFarlan and McKenney, the IT strategic impact grid can be used to detect potential benefits for businesses and it is based on four tenets: support, factory, turnaround and strategy [25]. Some researchers debate that big data have to be governed because of the data complexity according to the common principles considering the 6 V's characteristics, data privacy and security, heterogeneity, incompleteness, scale and timelines [26].

Recently, based on the explosion of the available data, pharmacogenomics has entered the big data era. Pharmacogenomics aims to improve healthcare based on individual genomic profiles. Alongside other factors that might affect drug responses, such as diet, age, diseases, lifestyle, environment and health state, pharmacogenomics has the potential to assist in the creation of individualized treatment plans for patients that leads to personalized medicine [32]. Similar to other areas of human genomics, pharmacogenomics is faced with data explosion of various omics data, such as genomics, transcrip-

tomics, metabolomics and proteomics data [31]. These huge amounts of heterogeneous data are usually obtained by using high-throughput technologies and they contain measurement errors and noise. Therefore, data preprocessing is required, as well as employing appropriate data mining techniques and high-performance computing.

Recent big data research in healthcare has paid attention to the possibility of establishing big data strategies for the medical and pharmaceutical industry, considering ontology engineering processes such as ontology enrichment, semantic workbench and finally enhancing data quality in the particular research domains [18]. Construction of an accurate predictive or descriptive model of biological processes and diseases, based on knowledge discovered from different sources, is just another idea for using big data in biomedical engineering [23].

In that regard, some researchers have classified the application of big data in health informatics mainly in three scenarios: high-risk patient management, risk of readmission prediction and mobile health services for lifestyle recommendations [28]. Other researchers have paid attention to creating biomedical ontology quality assurance considering a big data approach and scalable cloud computing environment with big data tools [27]. They highlighted the achieved speed-up of massively parallel algorithms using big data framework, which largely reduce the time required to perform exhaustive structural analysis of large ontological hierarchies and tracking versions of evolutional analysis.

Big data can also be used in other different manners: for dealing with the huge data flow and a tremendous amount of documents or to identify trends and predict future events [12]. Data analysis brings additional values to companies concerning time, information and knowledge, which brings operational intelligence [21, 29]. There are many different tools that provide huge benefits from using big data. These tools are sometimes associated with business pressures, but most often related to the legal pressures that companies are faced with as well as the need to manage a large amount of unstructured data and to search through them [4, 6, 7].

Taking into consideration the importance of gained knowledge from using big data concepts in many chemistry–associated sciences, the EU supports research efforts to many projects [40], providing a large amount of accessible data for researchers, including pharmaceutical, agrochemical, biotechnological, fragrances, and general chemical companies. Big data analysis methods are widely used in the analysis of chemical reactions and molecules from patients' data using pharmaceutical patterns and expert systems that provide scientific insights [38]. Toward creating a suitable framework for pharmacy's big data analysis, some researchers offer a platform for semantic exploitation of big data supported by a knowledge-based infrastructure for deep analysis, intended for outcome-oriented medicine, especially for cancer research [18].

The most used healthcare service applications usually use hybrid models including cloud computing and other services to manage big data in healthcare-related applications. This results in systematic innovations toward Industry 4.0 applications [41]. To provide some valuable analytics for supply chains as well as track-and-trace analytics in medicine and pharmacy, powerful database systems and in-process memory are required [39]. This concept requires a shift in computing architectures to handle data storage requirements and powerful server processing needed for analysis of a large volume of data in a secure manner.

Another related science is biomedicine that changes the landscape of privacy and personal information when personal health records are connecting to the patients' big omics data and clinical health data from electronic health records. This data integration plays a key role in personalized medicine [37]. Additionally, in the last decade besides biomedicine, exposome research is also widely used. Exposome data cover all exposure environmental factors from chemical and nonchemical agents, socio-behavioral, psychological factors such as stress, diet, endogenous and exogenous factors from the whole human lifespan [42].

Here the focus is put on predictive analytics and structured data intended to gain operational actionable intelligence [12]. For particular businesses, it is very important to deal with legal pressures and to manage a huge amount of unstructured data such as text, images and documents and to submit them for legal purposes [20].

One of the industry branches, whose executives were exposed to standardization and legal pressures to create large quantities of documents, is the pharmaceutical industry. Faced with increasing legal demands as well as business pressures and competition, many pharmaceutical companies implement emerging software solutions that help them to handle these pressures and to follow the latest global trends in the industry [22].

3 Experiences and Methodology

In this paper, we have considered a pharmaceutical company located in Skopje as a case study and systematically gained data, information and knowledge toward big data analytics possibilities. This company has encountered the problems of standardization, legal and business pressures

Table 1: The business pressures, organizational responses with used solutions, products and resources in the company.

business pressures	IT Solution	product/DB	period of time	company's resources	sectors included	external experts' help
Fast drug registration and approval	DMS, R&DExpert, eMPDExpert, RIMExpert, PhVExpert	Documentum/ MyPharmaExpert Suite/ Oracle	April 2005 - April 2006, Rollout 2007-2010	50 employees	5 departments	Front – 6 consultants, End – 6 consultants 50% of working time
Lack of data for decision making and quality control – need for improved data quality: accuracy, consistency, completeness	ALKASAP, MM, PP, CO, FI, S&D modules	SAP / Oracle DB	July 2006 – July 2009, Rollout 2009-2011	100 employees	10 departments	Front – 10 consultants End – 10 consultants 50% of working time
Increased competition of the market	CRM	SQL server	January 2006 – July 2006, Rollout 2007-2010	5 teams with 10 employees and 100 customers	3 departments	Front – 4 consultants End – 4 consultants 30% of working time
Insufficient HR control	HRM, PD, TEM, TM, OM, PA, PD	SAP /Oracle	March 2014 – June 2014, Rollout 2015-2016	4 teams with 25 employees	6 departments	Front – 3 consultants End – 3 consultants 50% of working time
The need for timelines and efficient reports	Big Data Analytics and Visualization (BI), MyPharma IR	Crystal Report, MyPharma 360 Intelligence Reporting	2007-2014	50 employees, 5 IT employees	IT department + managers of all levels	Front-1 educator, 2 consultants, 4 weeks
Implementation of Serialization	GenCode application	GenCode Plant Server, P&V System	I.2012-2016 II. 2017-2018 III. 5.2018-6.2018	I.7 sectors II.10 sectors III.EU Hubs	Regulatory Assurance, Quality Assurance, Logistics, Production, Quality Control, Service and maintenance, IT&T	GenCode consultant, EU Hubs consultant
Automation of Labs systems and Quality control.	Labs system integrated with ALKASAP	Simens/ SAP/ Oracle DB	I.03.2018-03.2019	3 sectors	Quality control, Regulatory Assurance, Quality control, Service and Maintenance	Front – 3 consultants – 4 months X 1 week

over more than one decade. The company's managers and employees had experienced extended working hours, preparing many relevant documents, dealing with a huge amount of structured, semi-structured and unstructured data. They had to deal with a lot of products and materials, following the quality demands according to the EU regulation, standards [47], competition and law. So, the company managers decided to implement emerging IT tools to help them to solve their legal and business problems. Firstly, the company started to change all legacy software solutions and to integrate data to obtain information silos with improved data quality, accuracy, timeliness, consistency and completeness (SAP ERP solution). The emerging IT solutions with high-performance hardware and appropriate software tools were implemented, as tabled in Table 1.

The company had to deal with structured and unstructured data flows with increasing data volume over time. For that reason, the company had chosen one of the leading technology solutions at that time and the software service development was with strong support from the company managers [19].

3.1 DMS Documentum solution

As a result of demands for drug registration and approval by the Medicine Drug Agency [8], preparing enormous amounts of documents and different unstructured data over strongly defined workflow processes were needed. These processes have involved many entries/actors such as authors, reviewers and approvers. The desired solution for this problem was more focused on managing unstructured data, documents, workflows, procedures and other data flows, which demanded using NoSQL databases. These huge amounts of documents for the Drugs Agency led to the implementation of Enterprise Content Management (ECM) Documentum based system, a special solution for the pharmacy and pharmaceutical industry [3] and processes of drug registration and control, called myPharmaExpert Suite [9], as shown in Figure 1. This solution had specific usage only for creating, testing, reviewing and approving documents for drugs registration, as a long-term process [3]. Documentum was implemented using Oracle DB and it was able to create a large number of documents' pages providing virtualization of documents for the na-

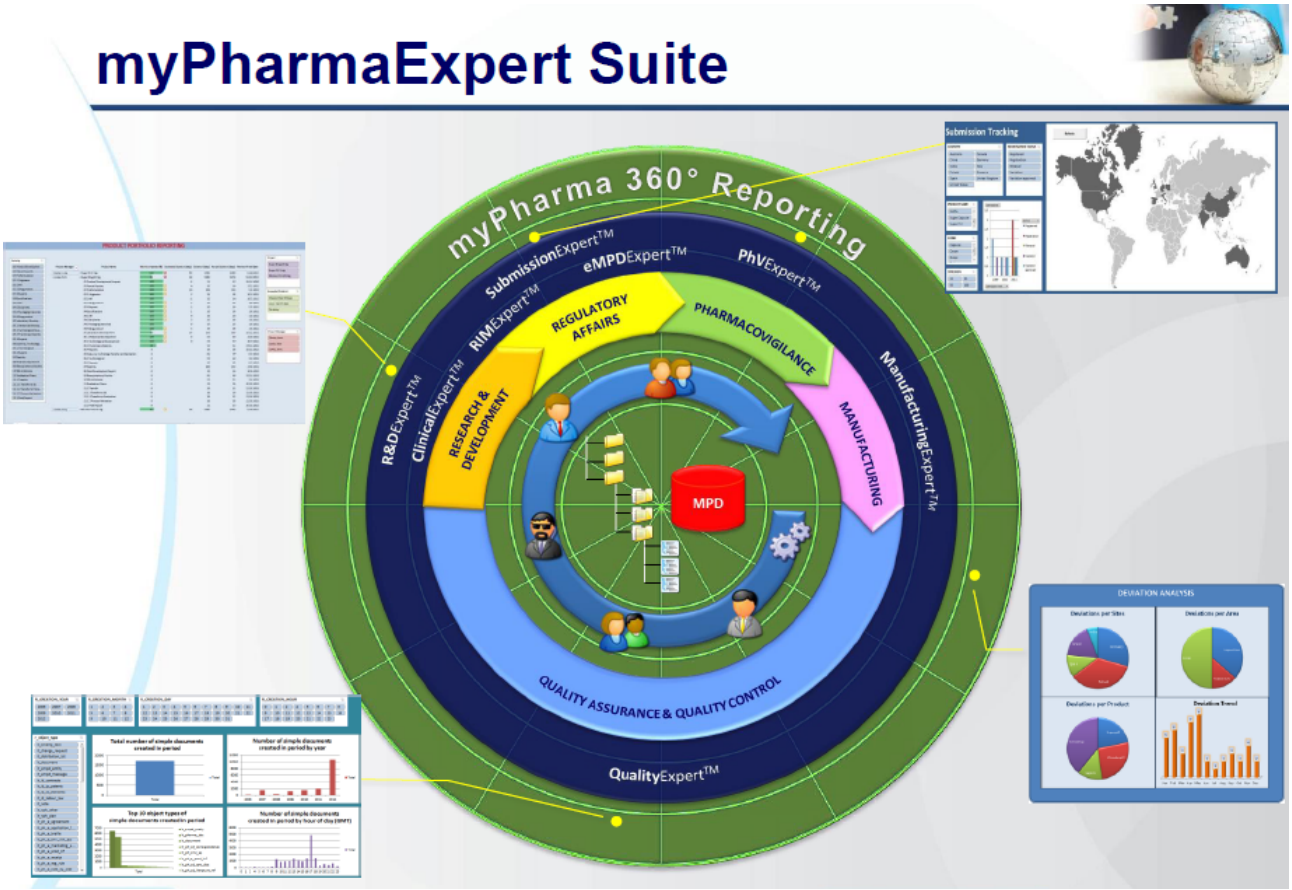

Figure 1: A Selected DMS Solution myPharmaExpert Suite based on Documentum [9].

tional medicine drug agency. The solution has helped the company to deal with a huge amount of unstructured data, occupying about 50 terabytes (TB) memory space and the same memory space for backup files. These data have been analyzed according to the big data concepts using text mining tools [3].

3 000-5 000 pages of documents for each drug were managed by the Documentum workflow management system with the possibility to manage many versions of the documents. All needed documents were prepared in a suitable format to create a virtual Dossier and then uploaded to the national drug agency web site. The process can take one to five years and it may have undergone several changes with many versions of some documents contained in the virtual Dossier. All corrected documents would have undergone the process of revision, approval and again reloaded on the drug agency web site, until the final registration of products and gaining a license for distribution, as shown in Figure 2.

The drugs produced by the company are subject to continuous revisions by the national authorities and the company's revision team. They had many versions that have to be stored, managed and reviewed. A large amount of these data flows connected with unstructured text and documents have to be stored for further usage. The company management and employees experienced positive effects of the implementation of myPharmaExpert Suite Documentum gaining virtual Dossier for the Drug agency and a lot of unstructured data for big data analysis.

3.2 ALKASAP Solution

According to the needs of data integration, all legacy in-house developed software solutions were migrated in SAP Enterprise Resource Planning (ERP) system known as SAP ERP. The solution was suitable for a lot of company problems, and its implementation was supported by the available consultants for ERP implementation. The development team has used JAD (Joint Application Development) methods for system analysis with previous set-up teams for implementation and succeed to implement all planned modules individually using external consultants' help, as shown in Table 1.

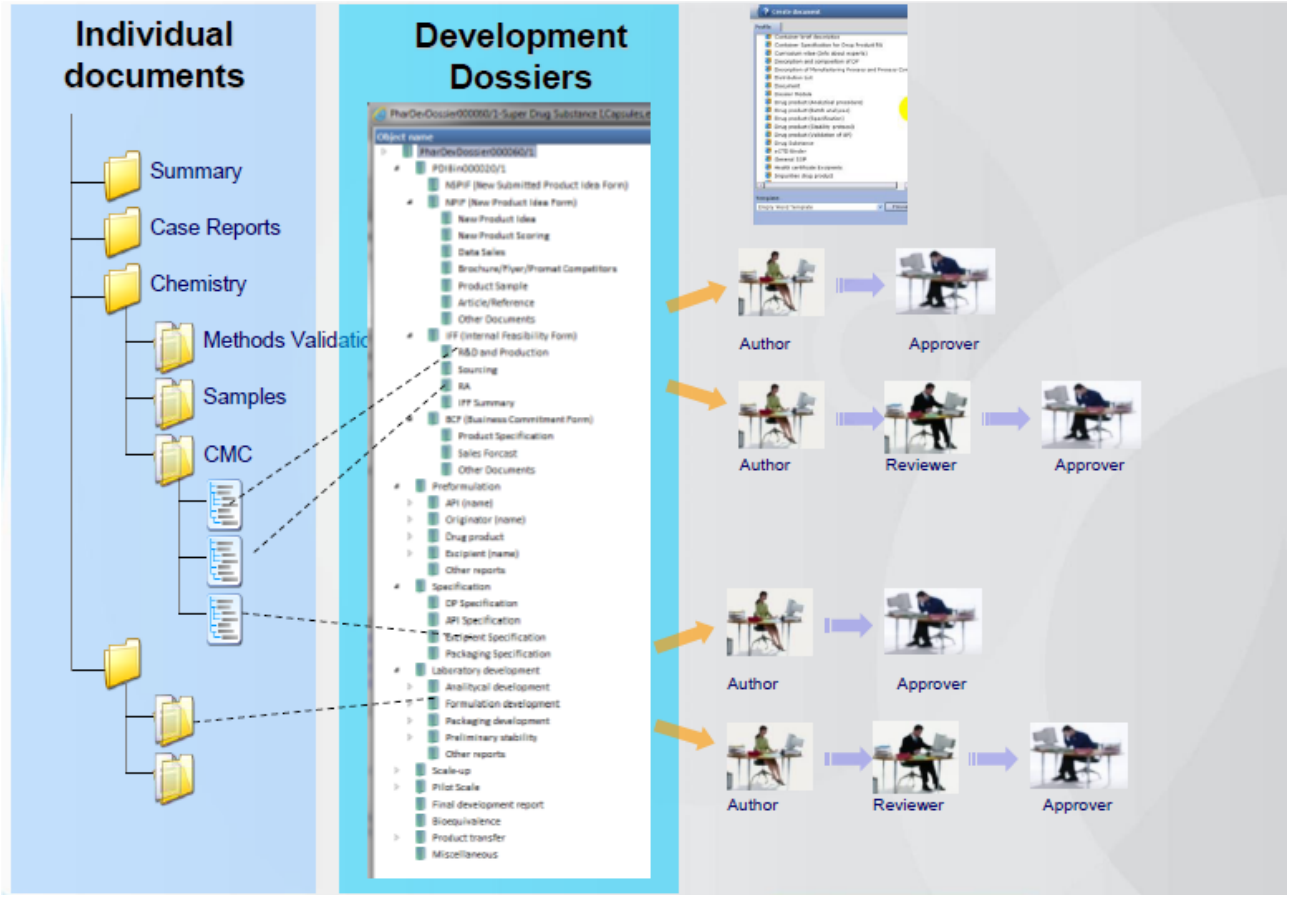

Figure 2: Processes of Dossier Development [9].

The data migration of the company's legacy databases was done using Oracle DB. Five SAP ERP modules were successfully implemented, internally popularly named as ALKASAP: Material/Warehouse management, Planning production, Sales and distribution, Business planning and Controlling and Finance module, as shown in Figure 3.

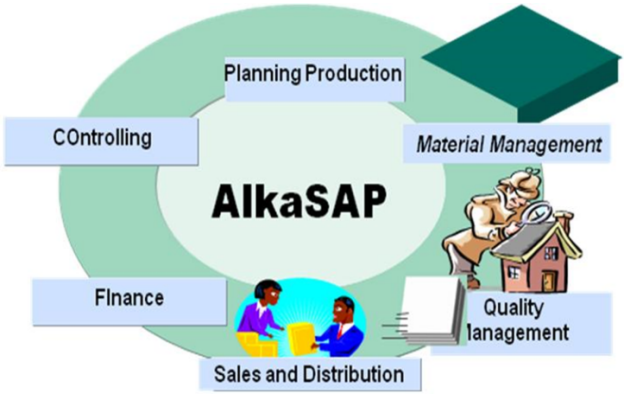

Figure 3: Implementation of ALKASAP in modules – First phase [1].

The implementation of needed modules of SAP ERP solution has taken almost 3 years [1]. It was just a structured part of the company's data used for data analysis following the big data concepts.

It was clear that this implementation reduced the time for planning and production for almost 4 000 products and more than 20 000 raw materials, increasing over time. Implementation has demanded a lot of time for operating activities that brought frustration and dissatisfaction among employees and line managers due to the highly increased workload. But for the company's managers, that meant getting temporal and online information and progress in the planning process, providing a clear view within 6 months. The implementation working teams were supported by SAP experts in the first stage.

3.3 Customer Relationship Management (CRM) Solution

The next part of the solution was the implementation of CRM systems particularly intended for the pharmaceutical industry's marketing, planning and control. This is a huge repository for big data analysis because of the permanent need for analyzing data associated with customer needs, preferences and opinions. It was a mobile solution that collected data from the visits through GPS (Global Positioning System) locations and tablet software for reporting [2]. This software has included an extensive range of unstructured data stored in a SQL database with unstructured parts, files and documents. Integration of ALKASAP with CRM has not been made, only some shared files were updated in CRM according to the defined workflow management procedures of ALKASAP. The IT staff and managers were aware that integration could be done by the implementation of a data warehouse with a suitable NoSQL database or by using appropriate big data analytics tools.

3.4 Implementation of HRM

Additionally, following the business needs, a Human Resources Management (HRM) system was implemented later on an SAP ERP solution. The purpose of this system was to follow the staff carriers, employees' data traceability according to the business logic with the modules Plan development and Training event management. This is also an excellent base for big data analysis taking into consideration the variety of voluminous data associated with employee carriers. The company has implemented the following modules: Travel management, Organizational management, Personal Administration and Personal development modules. This application has contained a large amount of data stored in Oracle DB [5], but many of them are stored as unstructured text files, pictures, movies and images and they have to be analyzed as big data [1].

The next stage included the rollout of the software tools and connection with the central IT building of four of fifteen subsidiaries abroad due to the high workloads and a large number of employees in these locations. Each ALKASAP implementation on remote locations requires four to five IT employees to work five to eight weeks. Delivery solutions were implemented in subsidiaries abroad and connected to the central database on the ALKASAP [5]. A big part of data was centralized, but some data were distributed, stored in company dislocated servers. Remote access was the prerequisite and a fundamental issue for the subsidiaries because of the real-time information and the quick company response.

Data security is one of the company's priorities and the highest standards of data protection were adopted. A large amount of data collected and stored in the company databases from ERP SAP (more than 50 TB in one location) are also stored on the Disaster Recovery Remote System. Backup is provided by disks' mirroring and covers all company data. The security issues are solved by the company's procedures for information security and processes of quality assurance, as shown in Figure 4.

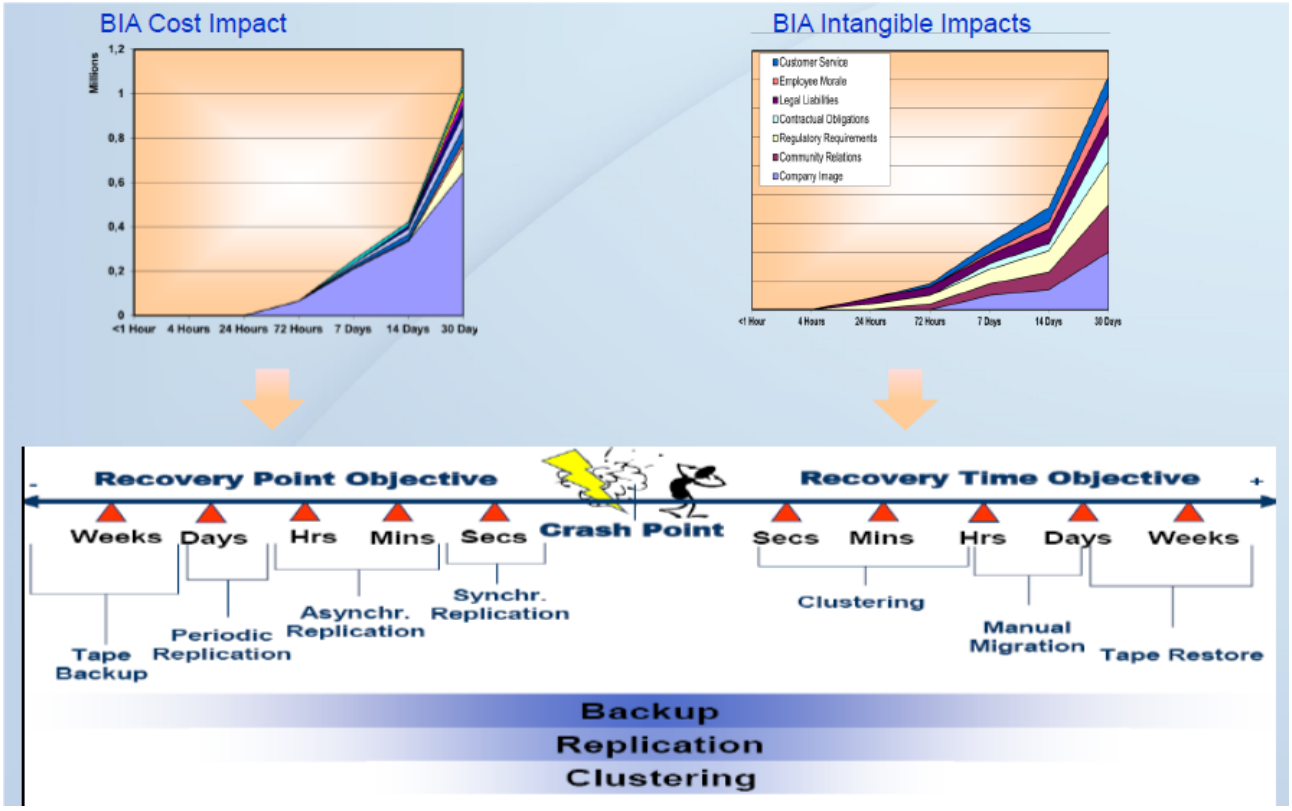

Figure 4: BIA and Information security solution with Disaster recovery [1].

3.5 Serialization

The serialization in the pharmaceutical industry was one of the most challenging activities [34]. The EU market demanded the implementation of a system for serialization for all products by February 2019. The concept of serialization uses big data and promises big changes for pharmaceutical companies. The company installed systems for serialization before the planned deadline and it has to deal with all challenges that address planned changes in the employed technology and the company's staff behavior. The legislation pressure transferred on the pharmaceutical companies by the EU directive 2016/161 in an endeavor to stop forgery from expanding into medical products [34]. Serialization is a process of assigning a unique code (*e.g.* GS1 2D) to every part of a commercial product to validate products and to enable entire quality control and for forgery detection [33]. Serialization enables identifying counterfeit products, as shown in Figure 5, and withdraws low-quality drugs quickly and efficiently, ensuring patient safety and hence avoiding the risk of negative publicity [34].

Led by serialization's demands, the company started with planning of the Serialization system implementation in September 2012, when a team of 7 strategic managers was set up. The team was planning to implement the serialization system, and then to connect to the EU Hubs and National Medical Verification Organization's (NMVO) systems. It was intended to verify all products, for each production level, as shown in Figure 6, taking into consideration device acquisition, their settings with the line controllers, connections with serialization line servers and company ERP or Manufacturing Execution System (MES).

The serial number provider with EPCIS (Electronic Product Code Information Services) repository had to be adjusted to the corporate IT solution, solving the specific activities for the company's serialization. Business processes connected with serialization were led by the serialization implementation teams (SIT) that had a wide responsibility for connecting all data processes with the basic system workflows, starting from SAP ERP Work orders to generation and control of serial numbers, as shown in Figure 7. This process has required a long time for staff education, especially SIT, project scheduling, equipment purchasing, setting up and implementing the whole project. But, at the end of the project implementation, the company had a great benefit and the EU directive 2016/161 was implemented. The task was completed and the company was connected with EU hubs and NMVO's systems.

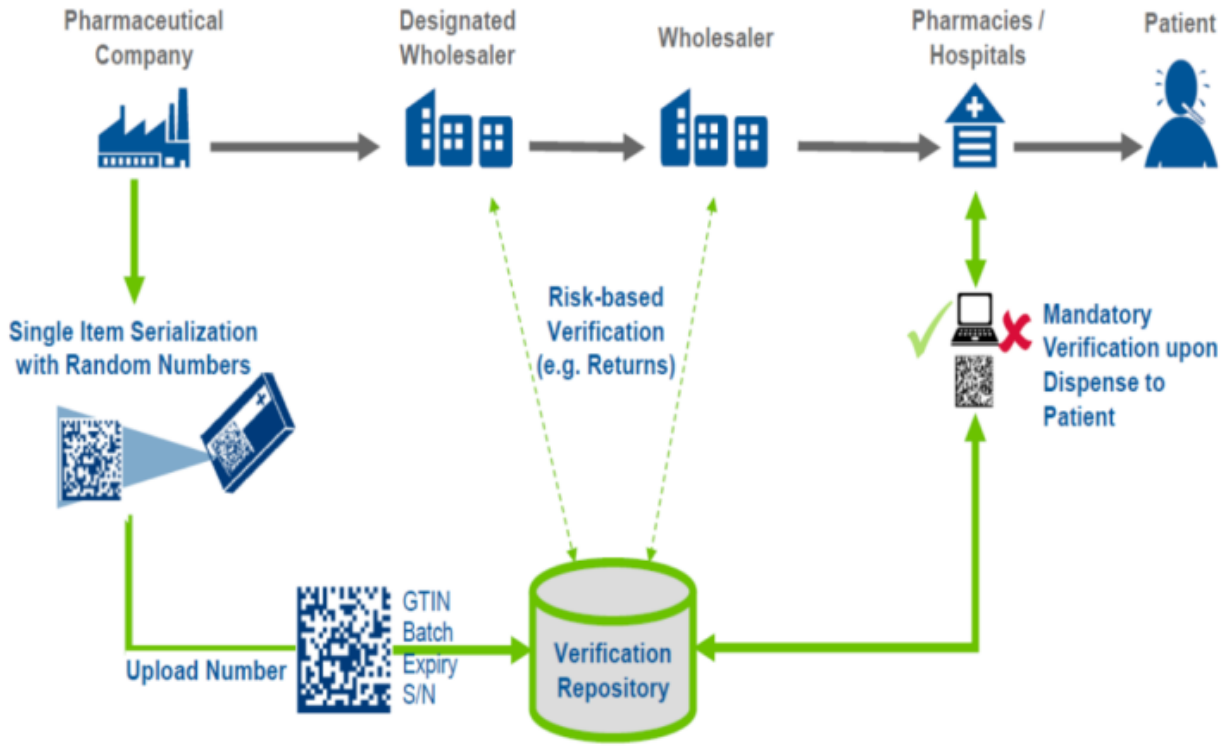

Figure 5: Identifying counterfeit pharmaceutical products.

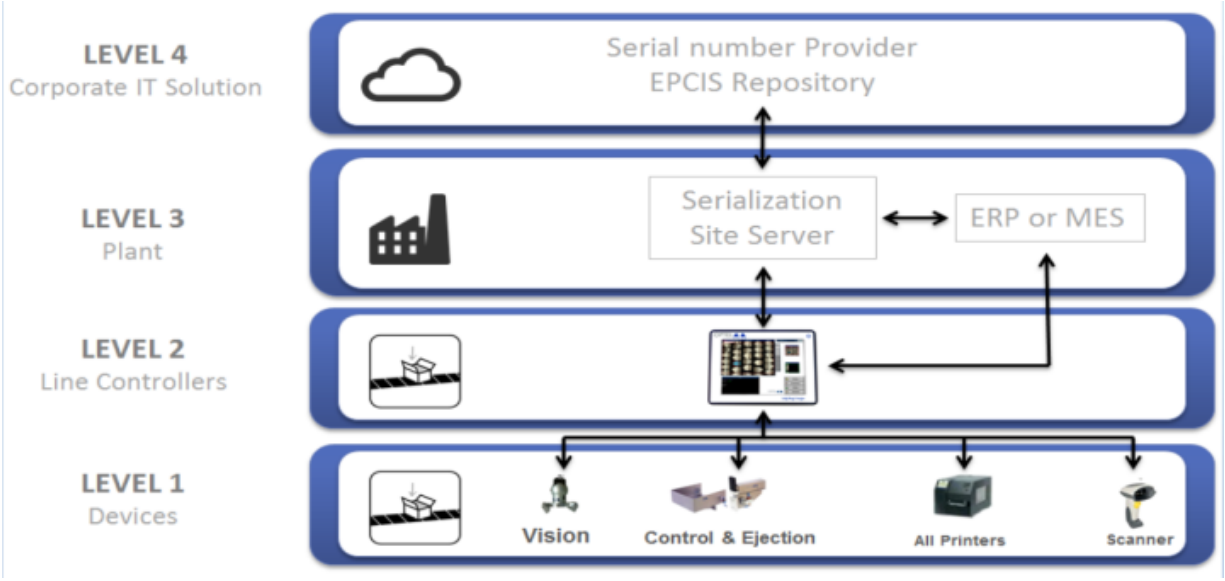

Figure 6: Implementation of a serialization system for each level.

Because the USC (Uniform System of Classification) code is a twenty-digit number that has to be printed on each product and each company package, the project included many ultrafast USC printers, scanners and cameras for code reading and spacers for unsuitable products. The EU USC server hubs control all drug distributions. The packages of hundreds or thousands of packs also have to be serialized with twenty digit USC codes. The tracking of each package is performed by connection to the EU server hubs and the sale had to be enabled only with medical practitioners' prescription and available unsold drugs with specific USC [1].

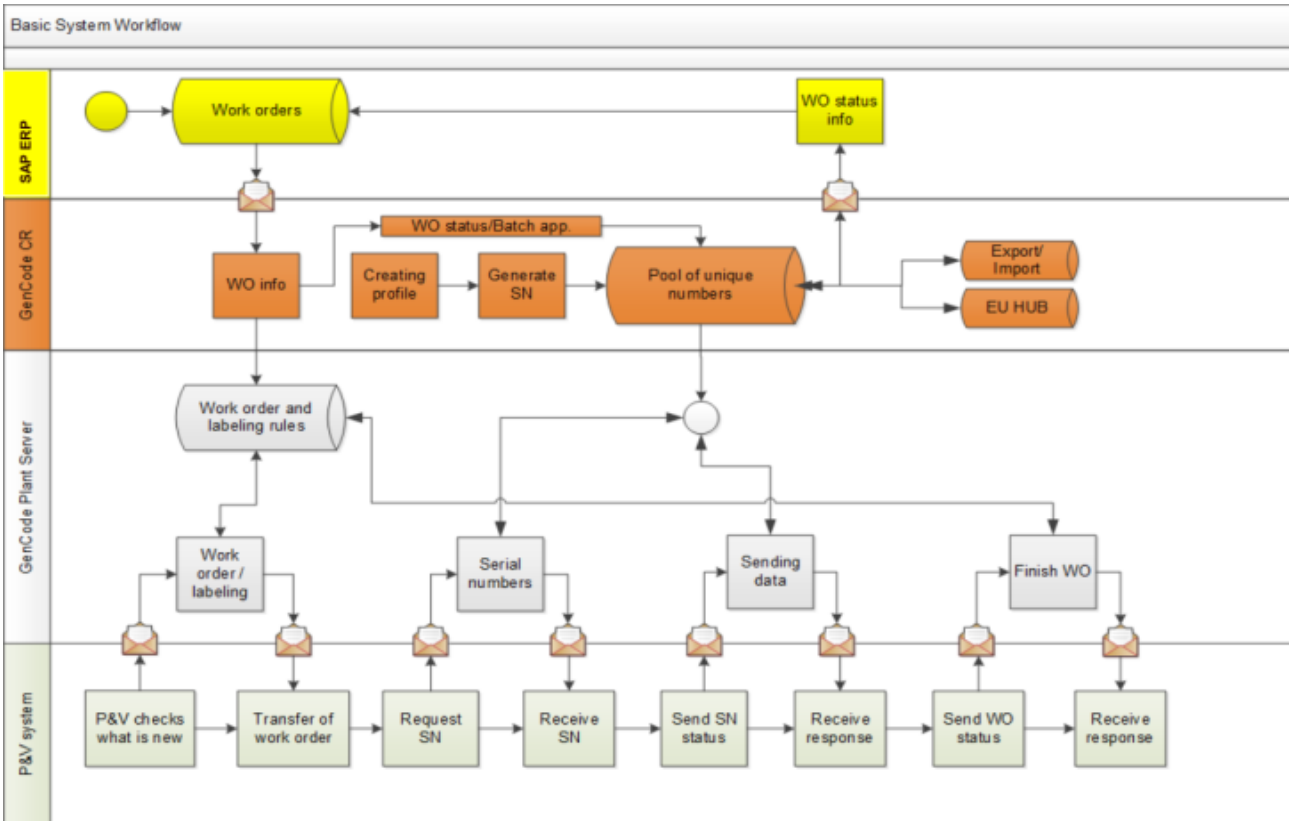

Figure 7: Data processes with basic system workflows.

3.6 Quality control

Providing quality of products and connected procedures has forced company managers to implement Quality Control (QC) system, Research and Development (R&D) system and Laboratory control system, collecting data from the company's measurement instruments. The research team in March 2018 started with the implementation of ALKASAP QC system, R&D and Labs data acquisition system [1].

Due to the importance of the QC, a separate building was built within the company, providing conditions for the QC and laboratory measuring devices. 300 cameras were installed during the final stage of implementation. They had produced video streaming data that was stored for further analysis. Also, laboratory measurement devices were connected with an automated data acquisition system and stored in ALKASAP databases, as the next step of the QC system implementation.

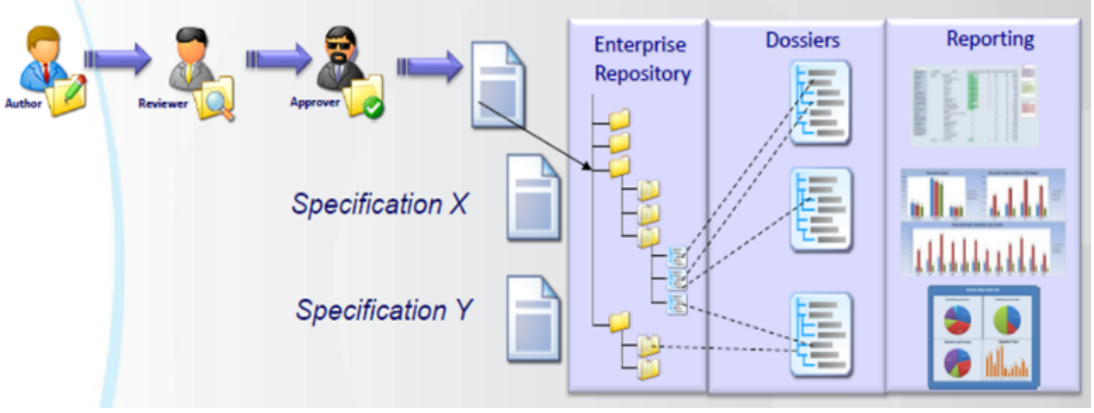

Figure 8: Processes of compliance with reporting possibilities [9].

The implementation had covered approximately 1 200 quality assurance standards that cover operational procedures for receiving materials, authorization and dissemination of all documents needed for the quality assurance in all product lines and for all drugs produced by the company. The company data was stored in an Oracle DB [16] and unstructured NoSQL databases for the possibility of analysis using dashboards [15]. The IT team members used Crystal Reports for data analysis and visualization, shown in Figure 8. Although the IT team consisted of 21 employees, their skills and knowledge in Oracle DB, ABAP (Advanced Business Application Programming) language, Java, Crystal Reports and other software tools, have covered all activities in support of ALKASAP, Documentum, CRM system and Business Intelligence with Crystal Reports [16]. The biggest part of business operations with outsourcing solutions has collected huge amounts of data stored in the data repository and was managed by highly trained people.

4 Discussion

The project of integration of all applications and databases was planned but the steering committee had not yet made a decision. The company had considered some options to integrate the Documentum solution used as myPharmaExpert with ALKASAP ERP, which is shown in Figure 9, as well as other data analytics and visualization tools to work on project integration.

However, new IT technology trends changed in the last decade toward using big data analytics and bioinformatics tools and methods that have influenced the company decision about system integration. The acquisition of Crystal Report by SAP brought Business Objects (BO) as the main data analysis and visualization tool. The company is planning also to roll out the implemented solution in their five subsidy companies abroad during the first quarter in 2020.

Although the pharmaceutical industry has started to analyze applications of big data, activities are still in their infancy [35]. The nature of the used data is biochemical and pharmacological including their regulatory mechanisms and targets if they are available [36]. However, the gained data in the pharmaceutical companies' databases, depending on applications, have different omics features, such as pharmacogenomics, genomics, transcriptomics, proteomics, metabolomics, exposomics, as shown in Table 2. These omics data should be the basis for further development of suitable pharmaceutical big data analytics applications to improve insights for the quality of the pharmaceutical and healthcare products as well as to perform meaningful data research using data mining tools and prediction analysis tools. This concept demands

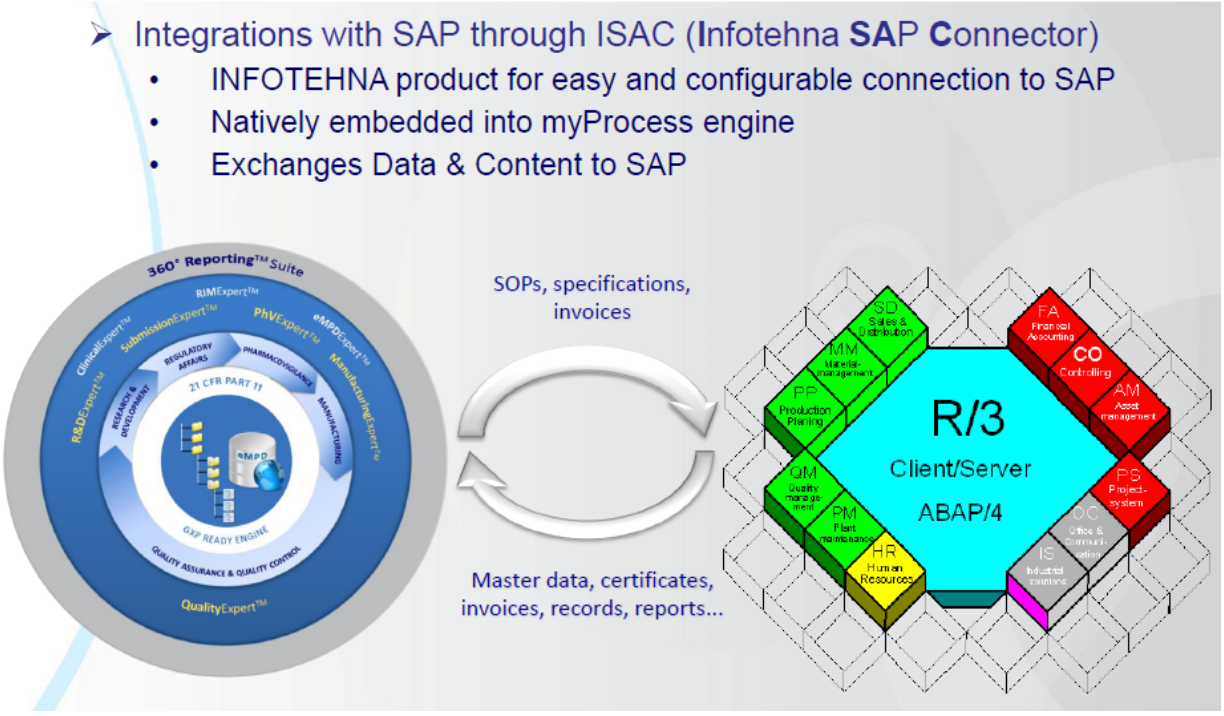

Figure 9: Possible integration of Documentum solution myPharmaExpert with ALKASAP [9].

Table 2: Opportunities for application of big data analytics in different companies.

company's application	use of –omics datasets and their analysis
DMS, R&DExpert, eMPDExpert, RIMExpert, PhVExpert	pharmacogenomics, genomics, transcriptomics, proteomics, metabolomics,/ data mining tools
ALKASAP, MM, PP, CO, FI, S&D modules	quality control and quality assurance to inspect the drug effects, as well as potential side effects as a part of exposome data
CRM with SQL server	personalized/precision medicine, predictions genomics, exposome
HRM on SAP ERP, SAP /Oracle	links of different modules (module for financial accounting, controlling, asset management, materials management, sales and distribution, human resources, etc.) with cross-application technologies, services and tools to systematize business processes and company's workflows
Big Data Analytics – Business Objects (BO), MyPharma 360 Intelligence Reporting, SAP data mining tools, omics data, personalization with PHR	enzymes, vitamins, and other pharmacogenomics results as drug metabolism and drug responses important for clinicians and healthcare providers
GenCode application, GenCode Plant Server, P&V System	quality assurance of omics technologies

the fastest database systems, so the company needs to move their business processes to an in-process-memory database, such as SAP HANA.

CRM system also has to be upgraded toward integration of big data and mobile devices according to the concept of the Internet of Things (IoT), considering the possibility of tracing employees' movement and using sensors for some data connected with their obligations and data that are provided.

Nowadays, the domain of integrative bioinformatics achieves very promising results and some insights have to be used for analysis by the company to highlight the important aspects of data integration. The huge amount of the company's data is suitable for the use of this concept.

5 Conclusion and further work

In this paper, we present a case study of a pharmaceutical company that implemented emerging IT solutions toward big data concepts to deal with permanent business, legal and market demands. The selected solutions increased the company's effectiveness and efficiency in creating, storing, managing, analyzing data and transforming them into information and organizational knowledge. The needs for creating, storing, monitoring and analyzing huge amounts of structured, semi-structured and unstructured data in different formats (*e.g.* text files, various documents, streaming data, workflow management data, pictures, movies, video and simulation data) require using big data analyses and reengineering processes.

The demands for efficient spending of working resources and their planning in the pharmaceutical company led to the implementation of five modules of SAP ERP system in the first phase. The next phase was the implementation of these modules in their subsidiaries abroad. This phase also included the implementation of HRM (2014) and QM (2015) with R&D processes. Marketing planning and controls were implemented on a separate application for marketing employees, aimed to work on tablets with GPS systems [2], providing online travel reports related to the batch processing with MS SQL server, connected with Oracle Database through common files for customers and products.

The quality assurance is connected to the serialization system. The company IT staff and the steering committee are aware that data integration was the problem that needed to be solved [1, 16]. So, they changed the IT development according to new trends in big data analysis. They are conscious that permanent investments in hardware, software solutions and licenses are an essential aspect of gaining competitive advantage [30].

The pharmaceutical companies' activities toward the application of big data analytics are still in an early stage, using tools for data analysis as Business Objects, Tableau and Tibco software [44, 45] and particular bioinformatics tools, such as EMBL-EBI, DisGeNET and UGENE, surveyed in [43]. Such biochemical and pharmacological data, which include the regulatory mechanisms and targets as

well as various omics data, should be essential for further investment in the development of suitable pharmaceutical big data analytics applications. These investments should improve the quality of healthcare and pharmaceutical products and additionally provide meaningful insights into the drug development.

Big data analytics demand in-memory databases, so the pharmaceutical company should migrate to a database such as SAP S/4HANA. The company should choose appropriate omics data analytics tools and data mining tools, which enable an analysis of various omics data.

As directions for further work, we suggest using IoT measuring devices that measure environmental conditions and ecological parameters to obtain GIS data. These data should be integrated with CRM data. Also, serialization data, as well as laboratory data, can be beneficial for patients' personal health records (PHR). These data can be used in metabolomics, exposomics data, as the influence of drugs on each patient, allergomics data or data about patients that belong to groups with particular characteristics. This can be achieved by using suitable integration of PHR data and appropriate integrative bioinformatics tools for big omics data analysis.

References

[1] ALKASAP (2016), Alkaloid AD Skopje, http://www.alkaloid.com.mk/alkasap-en.nspx, 20.5.2016.

[2] ACRM (2016), Alkaloid CRM system, http://www.alkaloid.com.mk/alkasap.nspx, 20.5.2016.

[3] ASC (2016), Alkaloid Standard and control, http://www.alkaloid.com.mk/standards-and-controls.nspx, 20.5.2016.

[4] Azarmi, B. (2016). Scalable Big Data Architecture. *A Practitioner's Guide to Choosing Relevant Big Data Architecture. Apress, Berkeley*.

[5] Berger C., Oracle Advanced Analytics 12c & SQLDEV/Oracle Data Miner 4.0, New Features, 2014

[6] Carter, K. B. (2014). *Actionable Intelligence: A Guide to Delivering Business Results with Big Data Fast!*. John Wiley & Sons.

[7] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4).

[8] Macedonian Agency for medicines and medicine devices, http://malmed.gov.mk/index.php/2015-07-11-10-47-48, 12.5.2016.

[9] MPES (2014), Platforms and Algorithms for Big Data Analytics, MyPharmaExpert Suite, Infotehna group LLC, http://cdn.ddfevent.com/pdf/SinisaBelina_2014.pdf, 14.2.2014

[10] Dupin-Bryant, P. A., & Olsen, D. H. (2014). Business intelligence, analytics and data visualization: A heat map project tutorial. *International Journal of Management & Information Systems (IJMIS)*, *18*(3), 185-200.

[11] Kessel, P. V. (2014). Big data changing the way businesses compete and operate. *Insights on governance, risk and compliance, EY*.

[12] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

[13] Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, *154*, 72-80.

[14] Kwon, O., Lee, N., & Shin, B. (2014). Data quality management, data usage experience and acquisition intention of big data analytics. *International journal of information management*, *34*(3), 387-394.

[15] NVP Survey (2014), Big Data Executive Survey 2014, An Update on the Progress of Big Data in the Large Corporate World, www.newvantage.com, 10.4.2016

[16] Oracle Inc., (2013) Oracle data sheet, Powerful Visual Analytics for the Entire Organization Oracle Business Intelligence 12c

[17] Reddy C.K. (2013), Platforms and Algorithms for Big Data Analytics, Wayne State university, http://dmkd.cs.vt.edu/TUTORIAL/Bigdata/, 12.5.2016

[18] Seebode, C., Ort, M., Regenbrecht, C., & Peuker, M. (2013, October). BIG DATA infrastructures for pharmaceutical research. In *2013 IEEE International Conference on Big Data* (pp. 59-63). IEEE.

[19] Economist (2014), The Economist, View from the C-suite, Who's big on BIG DATA, The Economist Intelligence Unit Limited 2014

[20] Vance J., (2014), Big Data 50, http://startup50.com/wp-content/uploads/2014/10/BD50_Startups_V1.1.pdf, 2.6.2016

[21] Stodder, D. (2013). Data visualization and discovery for better business decisions. *TDWI Research*.

[22] Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2016). The impact of big data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology*, *84*(1-4), 631-645.

[23] Janga, S. C., Zhu, D., Chen, J. Y., & Zaki, M. J. (2015). Knowledge Discovery Using Big Data in Biomedical Systems [Guest Editorial]. *IEEE/ACM transactions on computational biology and bioinformatics*, *12*(4), 726-728.

[24] Stojanovic, N., Dinic, M., & Stojanovic, L. (2015, October). Big data process analytics for continuous process improvement in manufacturing. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 1398-1407). IEEE.

[25] Jokonya, O. (2015, November). Towards a conceptual framework for big data adoption in organizations. In *2015 International Conference on Cloud Computing and Big Data (CCBD)* (pp. 153-160). IEEE.

[26] Kanchi, S., Sandilya, S., Ramkrishna, S., Manjrekar, S., & Vhadgar, A. (2015, August). Challenges and Solutions in Big Data Management–An Overview. In *2015 3rd International Conference on Future Internet of Things and Cloud* (pp. 418-426). IEEE.

[27] Cui, L., Tao, S., & Zhang, G. Q. (2016). Biomedical ontology quality assurance using a big data approach. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *10*(4), 41.

[28] Fang, R., Pouyanfar, S., Yang, Y., Chen, S. C., & Iyengar, S. S. (2016). Computational health informatics in the big data age: a survey. *ACM Computing Surveys (CSUR)*, *49*(1), 12.

[29] Gorton, I., & Klein, J. (2014). Distribution, data, deployment: Software architecture convergence in big data systems. *IEEE Software*, *32*(3), 78-85.

[30] Macedonian Stock Exchange, http://www.mse.mk/mk/symbol/ALK, 10.10.2016.
[31] Ristevski, B., & Chen, M. (2018). Big data analytics in medicine and healthcare. *Journal of integrative bioinformatics*, *15*(3).
[32] Fan, J., & Liu, H. (2013). Statistical analysis of big data on pharmacogenomics. *Advanced drug delivery reviews*, *65*(7), 987-1000.
[33] Cordon, C., Garcia-Milà, P., Vilarino, T. F., & Caballero, P. (2016). Serialization in the pharmaceutical industry. In *Strategy is Digital* (pp. 47-64). Springer, Cham.
[34] Horalek, J., & Sobeslav, V. (2017, August). Track & Trace System with Serialization Prototyping Methodology for Pharmaceutical Industry in EU. In *International Conference on Mobile Web and Information Systems* (pp. 177-186). Springer, Cham.
[35] Kumar, S., & Singh, M. (2018). Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics*, *2*(1), 48-57.
[36] Brown, N., Cambruzziet al., (2018). Big data in drug discovery. In *Progress in medicinal chemistry* (Vol. 57, pp. 277-356). Elsevier.
[37] Costa, F. F. (2014). Big data in biomedicine. *Drug discovery today*, *19*(4), 433-440.
[38] Schneider, N., Lowe, D. M., Sayle, R. A., Tarselli, M. A., & Landrum, G. A. (2016). Big data from pharmaceutical patents: a computational analysis of medicinal chemists' bread and butter. *Journal of medicinal chemistry*, *59*(9), 4385-4402.
[39] Chircu, A. M., Sultanow, E., & Chircu, F. C. (2014, June). Cloud computing for big data entrepreneurship in the supply chain: using SAP HANA for pharmaceutical track-and-trace analytics. In *2014 IEEE World Congress on Services* (pp. 450-451). IEEE.
[40] Tetko, I. V., Engkvist, O., Koch, U., Reymond, J. L., & Chen, H. (2016). BIGCHEM: challenges and opportunities for big data analysis in chemistry. *Molecular informatics*, *35*(11-12), 615-621.
[41] Elhoseny, M., Abdelaziz, A., Salama, A. S., Riad, A. M., Muhammad, K., & Sangaiah, A. K. (2018). A hybrid model of internet of things and cloud computing to manage big data in health services applications. *Future generation computer systems*, *86*, 1383-1394.
[42] Dennis, K. K., et al. (2016). Biomonitoring in the era of the exposome. *Environmental health perspectives*, *125*(4), 502-510.
[43] Ramharack, P., & Soliman, M. E. (2018). Bioinformatics-based tools in drug discovery: the cartography from single gene to integrative biological networks. *Drug discovery today*, *23*(9), 1658-1665.
[44] Sallam, R. L., Tapadinhas, J., Parenteau, J., Yuen, D., & Hostmann, B. (2014). Magic quadrant for business intelligence and analytics platforms. *Gartner RAS core research notes. Gartner, Stamford, CT*.
[45] Howson, C., Sallam, R. L., Richardson, J. L., Tapadinhas, J., Idoine, C. J., & Woodward, A. (2018). Magic quadrant for analytics and business intelligence platforms. *Retrieved Aug*, *16*, 2018.
[46] Farmabrend Nova, Association of Foreign Innovative Pharmaceuticals Manufacturers Representatives, Code of Farmabrend Nova's Conduct During Promotion of Drugs Issued Under a Prescription and Communication Towards the Healthcare Professionals, 06.2018, https://www.efpia.eu/media/412640/macedonia_national-code.pdf?fbclid=IwAR27HfxLgLUhXNgR4gM7PQ7iaUE0SSdNkPoRU7OYxoqOH2V22w0rQv0vkvI, accessed 20.2.2020.
[47] European Federation of Pharmaceutical Industries and Associations (EFPIA) https://www.efpia.eu/, accessed 20.2.2020.